


Gene set enrichment analysis (GSEA) of sex-biased genes

MO Meritxell Oliva BS Barbara E. Stranger DC Daniel Cotter

Updated date: Feb 2, 2021

 An abbreviated version of this protocol was published in Science in Sep 2020

The impact of sex on gene expression across human tissues

DOI: 10.1126/science.aba3066

Detailed protocol

For sex-biased genes, we performed gene set enrichment analysis (GSEA) for a total of 32,112 gene sets comprised of TF targets, cell-type markers, disease, drug, hormone and immunological processes, and others. This collection includes 496 gene sets representing hormone synthesis and activity, reproduction, and sex differentiation. The file containing the gene set collection is provided herein (all_non_tfea_chip_genesets.modif.entrez.gmt.zip). Gene ids correspond to entrez ids.

For each tissue, GSEA was performed as follows:

1. X-linked genes were excluded, given the large differences in sex-bias effect sizes between X-linked and autosomal genes.
2. For each tissue, we ranked genes by effect size and sex effect sign, creating ranked gene lists with female-biased genes at the top and male-biased genes at the bottom.
3. The function *fgseaMultilevel* implemented in the R package *fgsea* (Ref. 1) was used for GSEA. It is based on the adaptive multilevel splitting Monte Carlo approach, which enables exceeding the results of simple sampling and calculating arbitrarily small p-values.
4. Enrichment p-values were corrected for inflation, based on estimated genomic inflation factor calculated for each of the tested gene set categories.
5. P-values were corrected for multiple gene set testing and defined significantly enriched gene sets as those with Benjamini & Hochberg FDR ≤ 0.25 .
6. A GSEA cross-tissue meta analysis was performed with Fisher's combined probability test, which enables identification of biological processes and mechanisms partially or totally shared across tissues. To take into account the sign of the enrichment in a particular tissue when combining p-values, the most abundant enrichment sign across tissues for a particular gene set was quantified, and the inverse of the p-value for the discordant-in-sign tissues was calculated and utilized for meta-analysis. Significant gene sets were defined as those with cross-tissue meta-analysis FDR ≤ 0.05 (Bonferroni correction).

Details of the gene set collection and results of the GSEA analyses are provided in Supplementary Materials text and Tables S7-9 (Ref. 2).

Ref. 1 - A. A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv 060012 [preprint]. 20 June 2016.

Ref. 2 - M. Oliva, M. Muñoz-Aguirre, S. Kim-Hellmuth, V. Wucher, A. D. H. Gewirtz, D. J. Cotter, P. Parsana, S. Kasela, B. Balliu, A. Viñuela, S. E. Castel, P. Mohammadi, F. Aguet, Y. Zou, E. A. Khramtsova, A. D. Skol, D. Garrido-Martín, F. Reverter, A. Brown, P. Evans, E. R. Gamazon, A. Payne, R. Bonazzola, A. N. Barbeira, A. R. Hamel, A. MartinezPerez, J. M. Soria, GTEx Consortium, B. L. Pierce, M. Stephens, E. Eskin, E. T. Dermitzakis, A. V. Segrè, H. K. Im, B. E. Engelhardt, K. G. Ardlie, S. B. Montgomery, A. J. Battle, T. Lappalainen, R. Guigó, B. E. Stranger, The impact of sex on gene expression across human tissues. Science eaba3066 (2020).

Related files

 all_non_tfea_chip_genesets.modif.entrez.gmt.zip



How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Oliva, M., Stranger, B. and Cotter, D. (2021). Gene set enrichment analysis (GSEA) of sex-biased genes. Bio-protocol Preprint. bio-protocol.org/prep807.
2. Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S. E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E. A., Skol, A. D., Garrido-Martín, D., Reverter, F., Brown, A., Evans, P., Gamazon, E. R., Payne, A., Bonazzola, R., Barbeira, A. N., Hamel, A. R., Martinez-Perez, A., Soria, J. M., Pierce, B. L., Stephens, M., Eskin, E., Dermitzakis, E. T., Segrè, A. V., Im, H. K., Engelhardt, B. E., Ardlie, K. G., Montgomery, S. B., Battle, A. J., Lappalainen, T., Guigó, R. and Stranger, B. E. (2020). The impact of sex on gene expression across human tissues. Science 369(6509). DOI: [10.1126/science.aba3066](https://doi.org/10.1126/science.aba3066)

Copyright: Content may be subjected to copyright.

